Statistica Sinica **30** (2020), 1005-1025
doi:https://doi.org/10.5705/ss.202016.0142

COMPOSITE LIKELIHOOD INFERENCE UNDER BOUNDARY CONDITIONS

Jing Huang[1], Yang Ning[2], Yi Cai[3], Kung-Yee Liang[4] and Yong Chen[1]

[1]*The University of Pennsylvania,* [2]*Cornell University,*
[3]*AT&T Services Inc and* [4]*National Health Research Institutes*

Abstract: Often, when a data-generating process is too complex to specify fully, a standard likelihood-based inference is not available. However, a composite likelihood can provide an inference based on a partial specification of a data-generating process. Furthermore, its robustness to model specification and computational simplicity makes the composite likelihood method widely applicable. This study conducts a theoretical investigation of the composite likelihood ratio test (CLRT) when the parameters of interest may lie on the boundary of the parameter space. Our main result shows that the limiting distribution of the CLRT is equivalent to that of the likelihood ratio test of a normal mean problem, in which the restricted mean of a multivariate normal distribution is tested based on one observation from a multivariate normal distribution with an inverse Godambe information matrix. Furthermore, we illustrate our general theoretical result by applying it to a variety of examples. Lastly, our simulation results confirm that the limiting distribution of the CLRT performs well in finite samples.

Key words and phrases: Boundary condition, composite likelihood, hypothesis testing, likelihood ratio test.

1. Introduction

Likelihood-based inferences are commonly used to model complex data. In some applications, specifying the data-generating process fully is difficult, or is not preferred, owing to a lack of knowledge about the true model or computational challenges. One possible solution is to conduct an inference based on a *partially* specified model. Introduced by Lindsay (1988), the composite likelihood approach has drawn a great deal of attention, and is widely used in many areas, such as biomedical research (Heagerty and Lele (1998); Molenberghs and Verbeke (2005); Wellner and Zhang (2000); Henderson and Shimakura (2003); Guan (2006); He and Yi (2011)), statistical genetics (McVean et al. (2004); Myers et al. (2005); Larribe and Fearnhead (2011)), geostatistics (Vecchia (1988);

Nott and Rydén (1999); Stein, Chi and Welty (2004); Padoan, Ribatet and Sisson (2010)), finance (Bhat, Varin and Ferdous (2010); Varin and Vidoni (2008)), social science, and many others.

Specifically, a composite likelihood is constructed as the product of a set of low-dimensional marginal or conditional densities. This approach is especially useful for modeling correlated data with a complex or unknown dependency structure, or for reducing a complex likelihood function with a high computational cost to a much simpler function. For instance, when a working independence assumption is adopted, the composite likelihood is also called an independence likelihood (Chandler and Bate (2007)). Because the data-generating process is partially specified, a composite likelihood inference is more robust than the standard likelihood-based inference. For further detail on composite likelihood methods, refer to Varin, Reid and Firth (2011), and the references therein.

In standard hypothesis testing problems, regularity conditions require that the parameter values are interior points of the parameter space under the null hypothesis. This assumption guarantees that the composite score function is zero at the maximum composite likelihood estimate. This first-order condition means the composite likelihood ratio statistic is asymptotically a mixture of weighted χ_1^2 (Kent (1982); Molenberghs and Verbeke (2005)). In many important applications, however, the parameters of interest may lie on the boundary of the parameter space. This constraint can arise from the definition of the parameter space, or from previous knowledge on the possible range of the parameter. For example, such a problem is encountered in diagnostic systematic reviews, which evaluate diagnostic accuracy by pooling the sensitivity and specificity of a dichotomized diagnostic test from multiple studies (Chen et al. (2014, 2015)).

As acknowledged in the literature, ignoring the boundary constraints often leads to a substantial loss of power (Self and Liang (1987); Chen and Liang (2010)). However, theoretical results for composite likelihood-based inferences under boundary conditions have not been established. In this study, we aim to fill this gap by providing the limiting distributions of the composite likelihood ratio tests (CLRTs) under boundary conditions. Furthermore, we apply our general theoretical result to study the following three examples: (1) a stratified case-control study; (2) diagnostic systematic reviews; and (3) adverse event detection in medical reports. In our simulation studies, we found that the naive method of ignoring the boundary constraints is grossly conservative, resulting in up to 48% less power than that of the proposed test.

The rest of this paper is organized as follows. In Section 2, we define the

composite likelihood, and introduce applications that include boundary problems. In Section 3, we first provide several regularity conditions, and then derive the asymptotic distribution of the CLRT. Here, we also provide a detailed calculation of the limiting distribution when all parameters of interest are on the boundary. In addition, we consider situations in which a subset of the parameters of interest lie on the boundary. In Section 4, we revisit the examples in Section 2. In Section 5, we present our simulation studies. Lastly, Section 6 concludes the paper.

2. Notation and Examples

2.1. Composite likelihood

Let $f(x;\theta)$ be the probability density function of a multidimensional vector random variable X, indexed by a p-dimensional parameter $\theta = (\theta_1, \ldots, \theta_p)^T$, where θ belongs to the parameter space Ω, a subset of R^p. We assume model identifiability, such that distinct values of θ correspond to distinct probability distributions. Suppose N independent random variables $X_1, \ldots, X_N$ are observed from the model $f(x;\theta)$. Let $\{\mathscr{A}_1, \ldots, \mathscr{A}_K\}$ be a set of marginal or conditional events, with associated likelihoods $L_i(\theta;\mathscr{A}_k) \propto \text{pr}(X_i \in \mathscr{A}_k;\theta)$, where $k = 1, 2, \ldots, K$ and K is the number of events. Following Lindsay (1988), a composite log likelihood can be construct as

$$\ell_c(\theta) = \sum_{i=1}^{N}\sum_{k=1}^{K} \omega_k \log L_i(\theta;\mathscr{A}_k),$$

where ω_k, for $k = 1, \ldots, K$, are nonnegative weights associated with the likelihood $L_i(\theta;\mathscr{A}_k)$. In particular, in longitudinal data analyses, the composite likelihood can be constructed by pooling the marginal densities, without considering the correlation between repeated measurements. The use of this likelihood is studied by Chandler and Bate (2007). The maximum of the composite likelihood at the parameter value $\hat{\theta}_c$ is the maximum composite likelihood estimator. Denote the first two derivatives of $\ell_c(\theta)$ as $U_c(\theta)$ and $h_c(\theta)$, respectively. Because each component of the composite likelihood is a true likelihood, it carries important features of the ordinary likelihood; for example, Bartlett identities hold for each component

$$E\{S_{ik}(\theta)\} = 0, \quad \text{and} \quad E\{-h_{ik}(\theta)\} = E\{S_{ik}(\theta)\}^2, \quad \text{for} \quad k = 1, 2, \ldots, K,$$

where $S_{ik}(\theta)$ and $h_{ik}(\theta)$ denote the first two derivatives, respectively, of $\log L_i(\theta;\mathscr{A}_k)$. Because the composite score function $U_c(\theta)$ is a linear combination

of component score functions $S_{ik}(\theta)$, $U_c(\theta)$ is an unbiased estimating equation. Assume that we are interested in testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Standard regularity conditions assume that θ_0 is an interior point of the parameter space. Under some extra regularity conditions, the CLRT, $2\{\ell_c(\hat{\theta}_c) - \ell_c(\theta_0)\}$, converges in distribution to a mixture of independent χ^2_1 variables (Varin, Reid and Firth (2011)), where the weights are the eigenvalues of $H(\theta_0)G^{-1}(\theta_0)$, with

$$H(\theta) = -E\left\{\sum_{k=1}^{K}\omega_k h_{ik}(\theta)\right\}, \quad J(\theta) = E\left\{\sum_{k=1}^{K}\omega_k S_{ik}(\theta)\right\}\left\{\sum_{k=1}^{K}\omega_k S_{ik}(\theta)\right\}^T,$$

and $G(\theta_0) = H(\theta_0)J^{-1}(\theta_0)H(\theta_0)$.

However, in many real applications, θ_0 may lie on the boundary of the parameter space, which makes existing asymptotic results based on the CLRT invalid. In the following, we present three examples in which boundary problems are encountered.

2.2. Test for positive associations in stratified case-control studies with sparse data

The stratified case-control design is widely used in epidemiological and genetic studies. In particular, in the ith stratum, $x_{i1}, \ldots, x_{in_i}$ denote $p \times 1$ vectors of potential risk factors of n_i cases, and $x_{in_i+1}, \ldots, x_{iN_i}$ denote the potential risk factors of m_i controls, where $m_i = N_i - n_i$ and $i = 1, \ldots, K$. A logistic regression model that accounts for stratum-specific effects is

$$\text{logit pr}(y_{ij} = 1 \text{ in stratum } i | x_{ij}) = \alpha_i + \beta^T x_{ij},$$
$$i = 1, \ldots, K, \text{ and } \; j = 1, \ldots, N_i,$$

where $y_{ij} = 1$ if the jth subject in the ith stratum belongs to the case group, and $y_{ij} = 0$ otherwise. The coefficients β quantify the effects of risk factors x_{ij} on disease status y_{ij}. In the case that n_i and m_i are uniformly bounded, but the number of stratum $K \to \infty$, the maximum likelihood estimator is known to be inconsistent. One common technique to solve this problem is to use the composite likelihood method proposed by Liang (1987), which extended the well-known Mantel–Haenszel estimator to logistic regression models with multiple risk factors. Specifically, for the (j, l) case-control pair of subjects in the ith stratum $(j = 1, \ldots, n_i; l = n_i + 1, \ldots, N_i)$, the conditional probability that x_{ij} is from the case, given that one of x_{ij} and x_{il} is from the case and the other is from the control, can be calculated as

$$\text{pr}(y_{ij}=1, y_{il}=0 | y_{ij}+y_{il}=1, x_{ij}, x_{il}; \alpha_i, \beta) = \frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}}.$$

Thus a composite likelihood can be formulated by considering all $n_i m_i$ possible pairs within the ith stratum,

$$L_i(\beta) = \prod_{j=1}^{n_i} \prod_{l=n_i+1}^{N_i} \frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}}.$$

A composite likelihood that combines the data from all K strata for β is then constructed by assigning the weight w_i to $L_i(\beta)$:

$$\ell_c(\beta) = \sum_{i=1}^{K} w_i \log \{L_i(\beta)\} = \sum_{i=1}^{K} w_i \log \left\{ \prod_{j=1}^{n_i} \prod_{l=n_i+1}^{N_i} \frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}} \right\}.$$

Without loss of generality, we consider the weights $w_i = N_i^{-1}$, where the maximum composite likelihood estimator reduces to the Mantel–Haenszel estimator when only a binary covariate is considered (Liang (1987)). Suppose some of the risk factors are known to be positively associated with the occurrence of a disease. Then, testing the null hypothesis $H_0 : \beta = 0$ is a boundary problem.

2.3. Test for heterogeneity in diagnostic systematic reviews

Diagnostic systematic reviews are a vital step in evaluating the accuracy of a diagnostic test. For a dichotomized diagnostic test, this usually involves drawing and comparing the sensitivity (Se) and specificity (Sp) from multiple studies. However, pooling the data is not straightforward. First, the estimated Se and Sp are typically negatively correlated between studies (Reitsma et al. (2005)). Second, there may be substantial between-study heterogeneity in paired indices (Moses, Shapiro and Littenberg (1993); Irwig et al. (1995); Rutter and Gatsonis (1995)). Such heterogeneity may arise from differences in study population characteristics, variability of assessments, and other factors.

To account for these challenges, Chen et al. (2014, 2015) proposed a composite likelihood-based approach that is robust and computationally convenient, in which they construct a composite likelihood function using an independent working assumption between Se and Sp. More specifically, considering a diagnostic review of m studies, denote n_{i11}, n_{i00}, n_{i01}, and n_{i10} as the number of true positives, true negatives, false positives, and false negatives, respectively, for $i = 1, \ldots, m$. Let $n_{i1} = n_{i11} + n_{i10}$ and $n_{i0} = n_{i01} + n_{i00}$ be the number of diseased and healthy subjects, respectively, and Se_i and Sp_i be the study-specific

Se and Sp, respectively. Assume that

$$n_{i11}|(n_{i1}, Se_i) \sim \text{Binomial}(n_{i1}; Se_i), \qquad n_{i00}|(n_{i0}, Se_i) \sim \text{Binomial}(n_{i0}; Sp_i),$$

$$g(Se_i) = X_i^T \beta_1 + \mu_{i1}, \quad g(Sp_i) = Z_i^T \beta_2 + \mu_{i2},$$
$$\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} \sim \text{BN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right),$$

where $g(\cdot)$ is a known link function, such as the logit function; X_i and Z_i are vectors of study-level covariates, possibly overlapping, related to Se_i and Sp_i; τ_1^2 and τ_2^2 capture the between-study heterogeneity in Se and Sp, respectively; and ρ describes the correlation between the random effects Se_i and Sp_i in the transformed scale.

Ignoring the correlation between Se_i and Sp_i, a composite likelihood can be constructed by setting $\rho = 0$ in the above model:

$$\ell_c(\theta) = \sum_{i=1}^{m} \log \left\{ \int q(n_{i11}|n_{i1}; Se_i)\phi(Se_i; \beta_1, \tau_1^2) dSe_i \right.$$
$$\left. \times \int q(n_{i00}|n_{i0}; Sp_i)\phi(Sp_i; \beta_2, \tau_2^2) dSp_i \right\},$$

where $\theta = (\beta_1^T, \beta_2^T, \tau_1, \tau_2)^T$, $\phi(\cdot; \cdot, \cdot)$ is the probability density function of a univariate logit normal distribution, and $q(\cdot|\cdot; \cdot)$ is the probability mass function of a binomial distribution. In this model, testing for heterogeneity in Se and Sp across all studies is equivalent to testing $H_0 : \tau_1^2 = \tau_2^2 = 0$. This is a hypothesis testing problem with boundary constraints.

2.4. Signal detection of adverse event (AE) reporting rate

The Vaccine Adverse Event Reporting System (VAERS) is a national vaccine safety surveillance program that collects information about adverse events (AE) that occur after the administration of vaccines licensed for use in the United States (Chen et al. (1994); Niu, Erwin and Braun (2001); Shimabukuro et al. (2015)). The surveillance data are structured in a table format, with vaccine-AE combinations (i.e., a certain type of AE after a particular vaccination) as the column variable, and the reporting year as the row variable (Huang, Zalkikar and Tiwari (2011)). In total, we have I years and J vaccine-AE combinations. The table cell n_{ij} shows the number of events reported for the jth vaccine-AE combination during the ith year. The total number of reported cases for the ith year is the marginal total of the ith row, denoted as $n_{i.}$. The total number of jth vaccine-AE combinations across all years is the marginal total of the jth column,

denoted as $n_{\cdot j}$. Let $n_{\cdot\cdot}$ denote the total number of events.

To account for the large number of zero cells in our data n_{ij}, we consider the following zero-inflated Poisson model:

$$n_{ij} \mid n_{i\cdot}, p_{ij}, w_{ij} \begin{cases} = 0 & \text{with probability} \quad w_{ij}, \\ \sim \text{Poisson}(n_{i\cdot} p_{ij}) & \text{with probability} \quad 1 - w_{ij}, \end{cases}$$

where w_{ij} is the probability of observing a true zero in the ith year and jth vaccine-AE combination, and p_{ij} is the probability of reported cases in the ith year and jth vaccine-AE combination. In order to account for the heterogeneity in the reporting proportion for a fixed jth vaccine-AE combination through year 1990 to 2015 (the most recent reporting year), we further assume that

$$\text{logit}(p_{ij}) \sim \text{N}(\beta_{0j}, \tau_j^2), i = 1, \ldots, I,$$

where $\beta_{0j} = \text{logit}(p_{0j})$, p_{0j} is the overall reporting proportion for the jth vaccine-AE combination across all years and τ_j^2 represents the variation in the reporting proportion. Similarly, we assume the weight w_{ij} is defined as follows

$$\text{logit}(w_{ij}) \sim \text{N}(\alpha_{0j}, \delta_j^2), i = 1, \ldots, I,$$

where $\alpha_{0j} = \text{logit}(w_{0j})$, w_{0j} is the overall probability of observing a true zero across all years and δ_j^2 represents the variation in observing a true zero.

A composite likelihood of the jth vaccine-AE combination can be constructed by multiplying the marginal densities of n_{ij}, ignoring their correlations:

$$\ell_c(\theta) = \sum_{i=1}^{I} \log \left\{ \int_0^1 \int_0^1 p(n_{ij}|n_{i\cdot}; p_{ij}, w_{ij}) p(p_{ij}|p_{0j}, \tau_j^2) p(w_{ij}|w_{0j}, \delta_j^2) dp_{ij} dw_{ij} \right\},$$

where $\theta = (w_{0j}, \beta_{0j}, \tau_j^2, \delta_j^2)$. Testing whether the reporting rate is the same across all years is equivalent to testing the hypothesis $H_0 : \tau_j^2 = \delta_j^2 = 0$, which is a boundary problem.

2.5. Additional examples

Other examples of problems with boundary constraints in composite likelihood inferences include testing for spatial correlations in a Gaussian random field model in geostatistics applications (Guan (2006)), and testing for serially dependence in time serial data in panel studies (Wellner and Zhang (2000)).

In the next section, we give our main results. The above examples are revisited in Section 4.

3. Main Results

3.1. Regularity conditions

To derive the asymptotic results, we impose the following regularity conditions.

R1: The first two derivatives of $\ell_c(\theta)$, with respect to θ on the intersection of the neighborhoods of the true parameter value, θ_0, and Ω, exist and are continuous. If θ_0 is on the boundary of Ω, the derivatives of $\ell_c(\theta)$ are taken from the appropriate side.

R2: The parameter space Ω is compact, the function $E\{\ell_c(\theta)\}$ is continuous, and $\theta_0 = \arg\max_{\theta\in\Omega} E\{\ell_c(\theta)\}$ is unique. There exists a function $A(X)$ with finite expectation, such that $|\sum_{k=1}^{K} \omega_k \log L_i(\theta; \mathscr{A}_k)| \leq A(X_i)$, for any $\theta \in \Omega$.

R3: On the intersection of neighborhoods of θ_0 and Ω, for any $1 \leq j, k \leq p$, it holds that $|\sum_{k=1}^{K} \omega_k(\partial^2 \log L_i(\theta; \mathscr{A}_k))/(\partial\theta_j\partial\theta_k)| \leq C(X_i)$, where $C(X)$ is a function of X with finite expectation.

R4: The variability matrix, $J(\theta)$, exists and is positive-definite at θ_0.

R5: The sensitivity matrix, $H(\theta)$, exists and is positive-definite at θ_0.

When studying the limiting distribution of an estimator for θ_0, we can think of it as a local problem, because only the values of θ near θ_0 are relevant, asymptotically. Thus, we assume that the parameter space near θ_0 can be locally approximated by a cone. This approach was used by Chernoff (1954) and Self and Liang (1987) in the context of likelihood ratio tests.

Definition 1. *The set $\Omega \subset R^p$ is approximated at θ_0 by a cone with a vertex at θ_0, C_Ω, if*

(1) $\inf_{x\in C_\Omega} \| x - y \| = o(\| y - \theta_0 \|)$ *for all* $y \in \Omega$

and

(2) $\inf_{y\in\Omega} \| x - y \| = o(\| x - \theta_0 \|)$, *for all* $x \in C_\Omega$.

We assume that the parameter space Ω is regular enough to be approximated by a cone with a vertex at θ_0, which is mild enough to encompass a wide variety of shapes for Ω. In the sections that follow, we first give the asymptotic distribution

of the CLRT statistic when θ_0 is on the boundary of Ω. Then, we derive the exact form of the asymptotic distribution in several cases.

For notational simplicity, rewrite the hypothesis $H_0 : \theta = \theta_0$ as $H_0 : \theta \in \Omega_0$, where $\Omega_0 = \{\theta : \theta = \theta_0 \in \Omega\}$. The complement of Ω_0 in Ω is denoted by Ω_1. For any subset of R^p, φ, we define $L_\varphi = \sup_{\theta\in\varphi} \ell_c(\theta)$. We also define the maximum composite likelihood estimator in the parameter space φ, $\hat{\theta}_c^{(\varphi)}$, as that value of θ in the closure of φ that maximizes $\ell_c(\theta)$. The CLRT statistic can be written as

$$CLRT = -2(L_{\Omega_0} - L_\Omega).$$

3.2. Limiting distribution of the CLRT statistic

First, we establish the $\sqrt{N}$-consistency of the maximum composite likelihood estimator.

Lemma 1. *If the regularity conditions* R1–R5 *hold, then as* $N \to \infty$, $\hat{\theta}_c$ *converges to* θ_0 *in probability. Moreover,* $N^{1/2}(\hat{\theta}_c - \theta_0) = O_p(1)$.

A proof is given in Section S1 of the online Supplementary Material. Now, we derive the asymptotic distribution of the CLRT.

Theorem 1. *If the regularity conditions* R1–R5 *hold,* θ_0 *is a limiting point of both* Ω_0 *and* Ω_1*, and the sets* Ω_0 *and* Ω_1 *are approximated by nonempty cones* C_{Ω_0} *and* C_{Ω_1}*, respectively, then under the null hypothesis, the asymptotic distribution of the CLRT is the same as the distribution of the likelihood ratio test of* $\theta \in C_{\Omega_0}$ *against* $\theta \in C_{\Omega_1}$*, based on one observation from a population with a multivariate normal distribution with mean* θ_0 *and covariance matrix* $H(\theta_0)^{-1}J(\theta_0)H(\theta_0)^{-1}$*, while the covariance matrix is misspecified as* $H(\theta_0)^{-1}$ *in the likelihood ratio test.*

The proof is based on two approximations. First, Ω is approximated by C_Ω locally around θ_0. This is justified by the $\sqrt{N}$-consistency of $\hat{\theta}_c$ and the definition of the approximating cones, given previously. The second approximation follows a similar argument to that of Self and Liang (1987), who used a quadratic function to approximate the composite likelihood. Further technical details are available in Section S2 of the online Supplementary Material.

In some special cases, the sensitivity matrix may be equal to the variability matrix, for example, in partial likelihood inferences for censored data (Cox (1975)). In such cases, the misspecified covariance matrix $H(\theta_0)^{-1}J(\theta_0)H(\theta_0)^{-1}$ is equal to $J(\theta_0)^{-1}$, and Theorem 1 reduces to Theorem 3 of Self and Liang (1987). However, in general, the equality is not satisfied (Molenberghs and Verbeke (2005)). The purpose of Theorem 1 is to reduce the general problem of

computing the limiting distribution of the CLRT to a problem of computing the distribution of

$$Q_{C_{\Omega_0}}(Z) - Q_{C_\Omega}(Z), \tag{3.1}$$

where $Q_\varphi(Z) = \inf_{\theta\in\varphi}\{Z-(\theta-\theta_0)\}^T H(\theta_0)\{Z-(\theta-\theta_0)\}$, $C_\Omega = C_{\Omega_0}\bigcup C_{\Omega_1}$, and $Z \sim MVN(0, G^{-1}(\theta_0))$, where $G(\theta_0) = H(\theta_0)J(\theta_0)^{-1}H(\theta_0)$ is known as the Godambe information matrix.

The limiting distribution of the CLRT in equation (3.1) is still complicated. In the following, we focus on an important special case in which the representation given in equation (3.1) can be simplified further.

3.3. An important special case

Partition the parameter vector, θ, into two parameter sets $\theta^T = (\gamma^T, \eta^T)$, where γ denotes the parameters of interest that will be tested, and η denotes the nuisance parameters. Then, further partition the parameter vector into four coordinates. Here, we adopt the notation in Self and Liang (1987): $(p_{11}, p_{12}, p_{21}, p-p_{11}-p_{12}-p_{21})$, where the first p_{11} coordinates of θ represent the parameters of γ with true values on the boundary; the next p_{12} coordinates represent the parameters of γ with true values not on the boundary; the next p_{21} coordinates represent the first p_{21} components of η with true values on the boundary; and finally, the remaining $p-p_{11}-p_{12}-p_{21}$ coordinates represent the last $p-p_{11}-p_{12}-p_{21}$ parameters of η with true values not on the boundary. Note that $p_1 = p_{11}+p_{12}$ represents the dimension of γ, and $p-p_1$ represents the dimension of η. In the rest of this paper, we denote the boundary value as zero for ease of presentation. In the following, we consider a special case in which two parameters of interest are on the boundary, and the remaining $(p-2)$-dimensional nuisance parameters are not on the boundary; that is, $H_0 : \gamma = 0$, with $\gamma = (\gamma_1, \gamma_2)$, $\gamma_1 \geq 0$, and $\gamma_2 \geq 0$. All three examples discussed in Section 2 belong to this setting.

The parameter configuration is give by $(2, 0, 0, p-2)$. Then, $C_{\Omega_0} = \{0\}^2 \times R^{p-2}$ and $C_\Omega = [0, +\infty)^2 \times R^{p-2}$. Partitioning $Z^T = (Z_\gamma^T, Z_\eta^T)$ and $G(\theta_0)$ and $H(\theta_0)$ with respect to (γ, η) as

$$G(\theta_0) = \begin{pmatrix} G_{\gamma\gamma} & G_{\gamma\eta} \\ G_{\eta\gamma} & G_{\eta\eta} \end{pmatrix}, \qquad H(\theta_0) = \begin{pmatrix} H_{\gamma\gamma} & H_{\gamma\eta} \\ H_{\eta\gamma} & H_{\eta\eta} \end{pmatrix},$$

$$G_{\gamma|\eta} = G_{\gamma\gamma} - G_{\gamma\eta}G_{\eta\eta}^{-1}G_{\gamma\eta}^T, \qquad H_{\gamma|\eta} = H_{\gamma\gamma} - H_{\gamma\eta}H_{\eta\eta}^{-1}H_{\gamma\eta}^T,$$

with some algebra, equation (3.1) reduces to

$$Z_\gamma^T H_{\gamma|\eta} Z_\gamma - \inf_{\gamma\in[0,+\infty)^2} (Z_\gamma - \gamma)^T H_{\gamma|\eta} (Z_\gamma - \gamma).$$

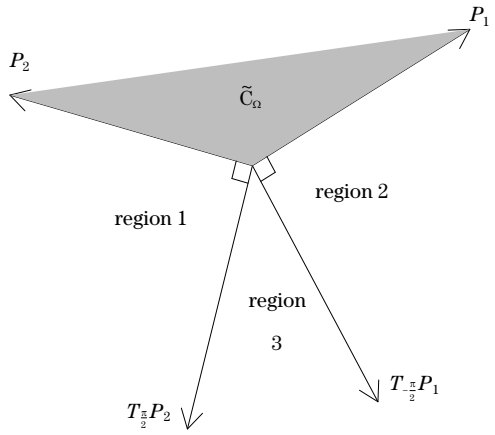

Figure 1. Diagram of the parameter space for the special case.

Let $H_{\gamma|\eta} = P^T P$, where P is a 2×2 nonsingular matrix, and denote $\tilde{C}_\gamma = \{\tilde{\gamma} : \tilde{\gamma} = P\gamma \text{ for any } \gamma \in [0, +\infty)^2\}$ and $\tilde{Z}_\gamma = PZ_\gamma$. Then, the limiting distribution of the CLRT can be rewritten as

$$\| \tilde{Z}_\gamma \|^2 - \inf_{\tilde{\gamma} \in \tilde{C}_\gamma} \| \tilde{Z}_\gamma - \tilde{\gamma} \|^2,$$

where $\| \cdot \|$ is the usual Euclidean metric. The calculation of the second term in the above equation depends on the location of $\tilde{Z}_\gamma$ relative to the boundary of $\tilde{C}_\gamma$. The shaded region in Fig. 1 represents $\tilde{C}_\gamma$, and (0, 0) is the origin. The angle in the shaded area is less than 180°. This is because that the convexity of C_γ is preserved under the linear mapping $\gamma \to P\gamma$. Denote the rotation matrix in a two-dimensional Euclidean space R^2 and the matrix $H_{\gamma|\eta}$ by

$$T_x = \begin{pmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{pmatrix}, \quad H_{\gamma|\eta} = \begin{pmatrix} a & b \\ b & d \end{pmatrix},$$

respectively. Further, denote the columns of P by P_1 and P_2, and the inner product of vectors a and b by $\langle a, b \rangle = b^T a$. It can be shown that the limiting distribution of the CLRT is given by

$$CLRT = \begin{cases} \| \tilde{Z}_\gamma \|^2 \sim U = \lambda_1 \chi_1^2 + \lambda_2 \chi_1^2 & \text{if } \tilde{Z} \text{ is in the shaded region,} \\ \left\{ \dfrac{\langle P_2, \tilde{Z}_\gamma \rangle}{\| P_2 \|} \right\}^2 \sim d^* \chi_1^2 / d & \text{if } \tilde{Z} \text{ is in region 1,} \\ \left\{ \dfrac{\langle P_1, \tilde{Z}_\gamma \rangle}{\| P_1 \|} \right\}^2 \sim a^* \chi_1^2 / a & \text{if } \tilde{Z} \text{ is in region 2,} \\ 0 & \text{if } \tilde{Z} \text{ is in region 3,} \end{cases}$$

where λ_1 and λ_2 are eigenvalues of $H_{\gamma|\eta}G^{-1}_{\gamma|\eta}$, and a^* and d^* are elements in the matrix $H_{\gamma|\eta}G^{-1}_{\gamma|\eta}H_{\gamma|\eta}$, such that

$$H_{\gamma|\eta}G^{-1}_{\gamma|\eta}H_{\gamma|\eta} = \begin{pmatrix} a^* & b^* \\ b^* & d^* \end{pmatrix}.$$

The mixing probabilities for the shaded region, region 1, and region 2 are

$$\pi_s = \cos^{-1}\frac{[((G_{\gamma|\eta})_{12})/(\{(G_{\gamma|\eta})_{11}(G_{\gamma|\eta})_{22}\}^{1/2})]}{2\pi},$$

$$\pi_1 = \cos^{-1}\frac{[((0,1)G_{\gamma|\eta}H^{-1}_{\gamma|\eta}P^TT_{\pi/2}P_2)/(\{(0,1)G_{\gamma|\eta}\binom{0}{1}\}^{1/2}\{P_2^TT^T_{\pi/2}PH^{-1}_{\gamma|\eta}G_{\gamma|\eta}H^{-1}_{\gamma|\eta}P^TT_{\pi/2}P_2\}^{1/2})]}{2\pi},$$

$$\pi_2 = \cos^{-1}\frac{[((1,0)G_{\gamma|\eta}H^{-1}_{\gamma|\eta}P^TT_{-\pi/2}P_1)/(\{(1,0)G_{\gamma|\eta}\binom{1}{0}\}^{1/2}\{P_1^TT^T_{-\pi/2}PH^{-1}_{\gamma|\eta}G_{\gamma|\eta}H^{-1}_{\gamma|\eta}P^TT_{-\pi/2}P_1\}^{1/2})]}{2\pi},$$

respectively. Thus, the asymptotic distribution of the CLRT is a mixture of U, $d^*\chi^2_1/d$, $a^*\chi^2_1/a$, and χ^2_0, with mixing probabilities π_s, π_1, π_2, and $1-\pi_s-\pi_1-\pi_2$, respectively.

4. Examples

4.1. Test for positive associations in stratified case-control studies with sparse data

Suppose the covariates x_{ij} are p-dimensional, and that two of them are known to be positively associated with the occurrence of a disease. To test the two positive associations simultaneously, the null hypothesis is $H_0 : \beta_1 = \beta_2 = 0$, and the alternative is H_a : any of $\beta_1, \beta_2 > 0$. In this case, the parameter vector, $\theta = (\beta_1, \beta_2, \ldots \beta_p)^T$, can be partitioned into two sets $\theta^T = (\gamma^T, \eta^T)$, for example, $\gamma = (\beta_1, \beta_2)^T$ and $\eta = (\beta_3, \ldots, \beta_p)^T$. The parameter configuration is given by $(2, 0, 0, p-2)$, $C_{\Omega_0} = \{0\}^2 \times R^{p-2}$, and $C_\Omega = [0, +\infty)^2 \times R^{p-2}$, which is the same as the special case discussed in Section 3.3. Thus, the asymptotic distribution of the CLRT is a mixture of U, $d^*\chi^2_1/d$, $a^*\chi^2_1/a$, and χ^2_0, with mixing probabilities π_s, π_1, π_2, and $1-\pi_s-\pi_1-\pi_2$, respectively, where the weights and mixing probabilities are calculated using the equations given in Section 3.3.

4.2. Test for heterogeneity in diagnostic systematic reviews

As discussed in Section 2.3, the null hypothesis is $H_0 : \tau_1^2 = \tau_2^2 = 0$, and the parameters involved are $\theta = (\tau_1^2, \tau_2^2, \beta_1, \beta_2)^T$. Partition this into two parameter sets $\theta^T = (\gamma^T, \eta^T)$, where $\gamma = (\tau_1^2, \tau_2^2)^T$ and $\eta = (\beta_1, \beta_2)^T$. The parameter configuration is given by (2, 0, 0, 2), $C_{\Omega_0} = \{0\}^2 \times R^2$, and $C_\Omega = [0, +\infty)^2 \times R^2$, which is the same as the special case in Section 3.3, with $p = 4$. Thus, the

limiting distribution of the CLRT is obtained by applying the results in Section 3.3.

4.3. Signal detection of AE reporting rate

To test whether the reporting rate is the same across all years for the jth vaccine-AE combination, the null hypothesis can be written as $H_0 : \tau_j^2 = \delta_j^2 = 0$. Similarly, we partition the parameter vector, $\theta = (\tau_j^2, \delta_j^2, \alpha_{0j}, \beta_{0j})^T$, into two parameter sets $\theta^T = (\gamma^T, \eta^T)$, where $\gamma = (\tau_j^2, \delta_j^2)^T$ and $\eta = (\alpha_{0j}, \beta_{0j})^T$. The parameter configuration is given by (2, 0, 0, 2), $C_{\Omega_0} = \{0\}^2 \times R^2$, and $C_\Omega = [0, +\infty)^2 \times R^2$, which is the same as the special case in Section 3.3, with $p = 4$. Thus, the results in Section 3.3 can be applied.

5. Simulation

To explore the finite-sample performance of the theoretical findings, we conduct two simulation studies. In the first example, we consider the CLRT in stratified case-control studies, as discussed in Section 4.1, with $p = 3$. In the second example, we test the heterogeneity of the sensitivities and specificities between multiple studies in diagnostic systematic reviews, as discussed in Section 4.2.

5.1. Test for positive associations in stratified case-control studies with sparse data

We simulate the data using three continuous covariates, and simultaneously test $\beta_1 = \beta_2 = 0$, where β_1 and β_2 are known to be nonnegative. The nuisance parameter β_3 belongs to $(-\infty, +\infty)$. The three covariates are independently simulated from a standard normal distribution. To mimic the selection procedure of a case-control study, we simulate 1,000 subjects for each stratum, and selecte five cases and five controls. The number of strata varies from 25 to 200, and the simulation is repeated 5,000 times. We compare the type-I error and the power of the CLRT based on the derived asymptotic distribution with that based on the naive χ_2^2 distribution.

As shown in Table 1, the method based on the derived asymptotic distribution controls the type-I error very well in all scenarios, whereas the naive method based on the χ_2^2 distribution yields grossly conservative type-I errors based on 5,000 simulations. The power gain is also substantial (up to 48%) using the derived asymptotic distribution of the CLRT. Thus, the naive method is more

Table 1. Empirical rejection rates (%) in 5,000 simulations of the CLRT to test for two regression coefficients in a stratified case-control study, based on different numbers of strata K, strata sizes N, and effect sizes.

		$\alpha = 0.10$ Rejection (%)		$\alpha = 0.05$ Rejection (%)		$\alpha = 0.01$ Rejection (%)	
$(\beta_1, \beta_2, \beta_3)$	(K, N)	CLRT	Naive	CLRT	Naive	CLRT	Naive
(0, 0, 0.1)	(25, 10)	9.8	0.4	4.7	0.1	0.7	0.0
	(50, 10)	10.5	0.3	5.2	0.1	1.0	0.0
	(100, 10)	9.5	0.2	4.3	0.0	0.6	0.0
	(200, 10)	10.1	0.5	4.9	0.1	1.1	0.0
(0.1, 0, 0.1)	(25, 10)	25.9	2.4	15.7	0.6	4.4	0.0
	(50, 10)	35.1	4.0	22.6	1.4	7.4	0.1
	(100, 10)	51.4	9.6	37.5	4.1	15.3	0.5
	(200, 10)	73.4	25.2	60.9	13.1	35.5	2.2
(0.1, 0.1, 0.1)	(25, 10)	37.9	4.7	25.0	1.6	7.9	0.1
	(50, 10)	56.1	11.0	41.4	4.4	17.9	0.5
	(100, 10)	78.1	26.6	64.8	13.8	38.3	2.5
	(200, 10)	95.1	62.0	90.0	44.7	73.3	16.2
(0.2, 0.2, 0.2)	(25, 10)	76.2	24.5	62.9	12.7	32.7	2.0
	(50, 10)	94.1	58.8	88.9	40.9	69.4	14.4
	(100, 10)	99.8	93.0	99.3	84.8	96.3	57.3
	(200, 10)	100.0	100.0	100.0	99.8	100.0	97.6

conservative and less powerful than the method based on the asymptotic distribution of the CLRT.

5.2. Test for heterogeneity in diagnostic systematic reviews

As described in Sections 2.3 and 4.2, we test for the heterogeneity in the sensitivities and specificities between multiple studies in diagnostic systematic reviews. We assume the covariates X_i and Z_i are univariate and independently generated from a standard normal distribution. We set $g(\cdot)$ as the logit function, and set $\beta_1 = g(0.95)$, $\beta_2 = g(0.90)$, and $n_{i1} = n_{i0} = 50$. We simulate μ_{i1}, μ_{i2} from a bivariate normal distribution with $\rho = -0.8$, and $\tau_1^2 = \tau_2^2$ take increasing values from 0, 0.05, 0.15, to 0.25. Then, n_{i11} and n_{i00} are simulated from the binomial distribution, as described in Section 2.3. We compare the type-I error and power of the CLRT based on the derived asymptotic distribution with that based on the naive χ_2^2 distribution under different numbers of studies m.

As shown in Table 2, the method based on the derived asymptotic distribution controls the type-I error reasonably well at nominal levels of 0.1, 0.05, and 0.01 in all scenarios. However, the naive method yields grossly conservative type

Table 2. Empirical rejection rates (%) for 5,000 simulations using the CLRT for heterogeneity in diagnostic systematic reviews, based on different numbers of studies m.

		$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		Rejection (%)		Rejection (%)		Rejection (%)	
$\tau_1^2 = \tau_2^2$	m	CLRT	Naive	CLRT	Naive	CLRT	Naive
0	10	0.6	0.0	0.0	0.0	0.0	0.0
	20	2.5	0.3	0.7	0.1	0.0	0.0
	40	4.3	1.3	1.8	0.2	0.1	0.0
	80	7.0	2.3	3.0	0.8	0.3	0.0
	160	9.3	3.2	4.4	1.4	0.6	0.1
0.05	10	33.4	20.3	22.9	12.5	8.8	4.5
	20	33.9	20.8	23.1	13.4	9.1	4.8
	40	47.1	30.2	34.2	20.1	14.6	8.14
	80	57.8	39.7	44.2	29.3	21.7	12.5
	160	74.9	58.6	62.7	47.2	37.5	25.5
0.15	10	52.5	36.2	39.9	26.8	20.9	13.8
	20	68.7	55.6	59.0	46.6	39.2	28.7
	40	71.0	55.8	59.1	46.8	40.7	31.2
	80	97.7	95.7	96.2	93.6	91.6	87.3
	160	100.0	100.0	100.0	99.9	99.8	99.6
0.25	10	70.6	58.6	62.1	50.3	44.1	34.3
	20	92.2	87.2	88.8	82.8	78.2	70.3
	40	97.9	96.6	96.9	95.1	93.6	90.4
	80	100.0	99.8	99.9	99.8	99.7	99.3
	160	100.0	100.0	100.0	100.0	100.0	100.0

I error in this example based on 5,000 simulations. The power gain is up to 15% as a result of using the asymptotic distribution of the CLRT.

6. Application to a Systematic Review of Modern Imaging Technologies for the Surveillance of Melanoma

Melanoma is a type of skin tumor that develops from pigment-containing cells known as melanocytes. Melanoma is a less common type of skin cancer, but is much more dangerous when not found early, resulting in the majority (75%) of deaths related to skin cancer (Lo and Fisher (2014)). Modern imaging technology can be used for the early detection of melanoma metastasis, and provides a cost-effective surveillance approach (Jemal et al. (2009)). Currently, the most commonly used diagnostic imaging technologies for melanoma include ultrasonography (US), computed tomography (CT), positron emission tomography (PET), and a combination of the latter two (PET-CT). In addition to evaluate

the relative performance of these contemporary diagnostic imaging technologies in diagnosing melanoma for patients at different stages, for example, regional and distant lesions, it is also important to quantify the heterogeneity in the imaging technologies, in terms of their operating characteristics, for example, the variability in sensitivity and specificity across study populations. Xing et al. (2011) conducted a diagnostic review based on 98 published studies of 10,528 patients, carried out between January 1, 1990, and June 30, 2009. The number of studies for each diagnostic imaging technology and type of cancer (regional and distant metastasis) are shown in Table S1 in the online Supplementary Material.

We applied the proposed composite likelihood model described in Section 2.3. Then, we used the proposed CLRT described in Section 3.3 to test for heterogeneity in the sensitivities and specificities of the imaging technologies for different cancer types across multiple studies for seven technology-cancer combinations. Our results show that, when diagnosing regional metastatic melanoma, all four imaging technologies have significantly heterogeneous sensitivity or specificity across multiple studies ($p < 0.001$). In diagnoses of distance metastatic melanoma, CT and PET have significant heterogeneity of sensitivity or specificity across studies ($p < 0.05$), whereas the combination of CT and PET does not ($p > 0.1$). Figure S1 in the online Supplementary Material shows the ranges of the sensitivities and specificities of the imaging technologies when diagnosing cancer types across studies. Note that the specificities are generally higher than the sensitivities, in general, and the heterogeneity of specificity is smaller than the sensitivity across studies for all diagnosis methods. The heterogeneities of both sensitivity and specificity are higher in the diagnosis of regional metastatic melanoma than they are for those of distant metastatic melanoma. We also compared the results of the proposed methods to the naive method, which ignores the boundary constraints. The naive method was more conservative, although it still identified the significant heterogeneity in the sensitivity or specificity of the four imaging technologies used to diagnose regional metastatic melanoma, and in the PET technology used to diagnose distant metastatic melanoma, given that the magnitudes of these heterogeneities are relatively large. However, the naive method failed to identify the significant heterogeneity in the sensitivity or specificity of the CT technology used to diagnose distant metastatic melanoma, which was identified by the proposed method.

7. Discussion

In this study, we derive the asymptotic distribution of the CLRT when a subset of the testing parameters lie on the boundary of the parameter space, following the work of Self and Liang (1987) and Chen and Liang (2010). The former work studied the asymptotic behavior of the regular likelihood ratio test when the parameters of interest lie on the boundary. The latter extended the results to deal with situations when a subset of the parameters of interest are tested based on the pseudolikelihood of Gong and Samaniego (1981). Considering that the composite likelihood approach has become increasingly popular, and no previous works have examined the asymptotic behaviors of the CLRT under boundary constraints, our work strives to fill this gap. Note that the results presented here are very broad, in that any partially specified models can be considered a special case of a composite likelihood.

The composite likelihood ratio-based inference under boundary constraints is a difficult question. Although the asymptotic results derived here yield well-controlled type-I errors and adequate power, calculating the test statistic becomes more complicated as the number of boundary parameters increases. Another alternative is to use numerical methods. However, caution is required in terms of validity, choice of tuning parameters, computational cost, and practical performance. For example, standard nonparametric and parametric bootstrap methods lead to inconsistent estimates when the parameter is on the boundary of the parameter space (Andrews (2000)). This inconsistency is due to the non-smoothness of the empirical distribution from which the bootstrap samples are generated. Andrews (2000) proposed subsampling and m-out-of-n bootstrap methods for obtaining consistent estimators of the limiting distributions of test statistics under boundary constraints. The m-out-of-n bootstrap provides a smoothing operation on the empirical distribution function, by resampling with replacement a smaller sample size m from the original n samples, where m is of a smaller order than n. However, these numerical methods require additional tuning parameters, and the performance varies in practice. For details of m-out-of-n bootstrap, please refer to Politis, Romano and Wolf (1999), and the references therein.

To examine its practical performance, we implemented the m-out-of-n bootstrap method to estimate the limiting distribution of the CLRT statistic and to test for positive associations in the example of stratified case-control studies with sparse data, as described in Section 5.1. We adopted the method proposed by Bickel and Sakov (2008) to make a data-adaptive choice on resampling size m,

with q = 0.85. The results are shown in Section S4 of the online Supplementary Material. Note that the performance of the m-out-of-n bootstrap method is between that of the proposed method and the naive method. More specifically, the empirical type-I error rates of the m-out-of-n bootstrap method are closer to the nominal levels than the naive method is, but are still very conservative compared with the proposed method. The power loss of the m-out-of-n bootstrap method was substantial, compared with the proposed method. In addition, the m-out-of-n bootstrap method required more computational hours than the naive and proposed methods did. Recently, Chen et al. (2018) proposed a a testing procedure that can better balance computational simplicity and statistical power for inference under boundary conditions using likelihoods where the second-order Bartlett identity does not hold. The proposed method was based on a conditional technique proposed by Susko (2013) and Bartholomew (1961). The test statistic converges weakly to a simple χ^2 distribution with data-dependent degrees of freedom, given the number of parameters lying on the boundary. It avoids the calculation of mixing proportions with limited sacrifice of statistical power.

Supplementary Material

In the online supplementary material, we provide detailed proofs for Lemma 1 and Theorem 1, and present an example that uses data on a diagnostic review of imaging technologies for the surveillance of melanoma, taken from Xing et al. (2011). We also provide the empirical type-I error rates and power of tests for positive associations in stratified case-control studies with sparse data, as described in Section 5.1. Here, we use a composite likelihood ratio statistic and its limiting distribution, calculated using the m-out-n bootstrap method.

Acknowledgments

This research was supported in part by the National Institutes of Health (NIH) under award numbers R01LM012607 R01AI130460, and R01HD099348, and by the National Science Foundation under award number DMS-1854637.

References

Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68**, 399–405.

Bartholomew D. J. (1961). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society. Series B (Methodological Methodology)*, 239—281.

Bhat, C. R., Varin, C. and Ferdous, N. (2010). A comparison of the maximum simulated like-

lihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. *Advances in Econometrics* **26**, 65.

Bickel, P. J. and Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* **18**, 967–985.

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence log-likelihood. *Biometrika* **94**, 167–183.

Chen, R. T., Rastogi, S. C., Mullen, J. R., Hayes, S. W., Cochi, S. L., Donlon, J. A. and Wassilak, S. G. (1994). The vaccine adverse event reporting system (vaers). *Vaccine* **12**, 542–550.

Chen, Y., Huang, J., Ning, Y., Liang, K.-Y. & Lindsay, B. G. (2018). A conditional composite likelihood ratio test with boundary constraints. *Biometrika* **105**, 225–232.

Chen, Y. and Liang, K.-Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97**, 603–620.

Chen, Y., Liu, Y., Ning, J., Cormier, J. and Chu, H. (2015). A hybrid model for combining case–control and cohort studies in systematic reviews of diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**, 469–489.

Chen, Y., Liu, Y., Ning, J., Nie, L., Zhu, H. and Chu, H. (2014). A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. *Statistical Methods in Medical Research*, 0962280214562146.

Chen, Y., Ning, J., Ning, Y., Liang, K.-Y. and Bandeen-Roche, K. (2017). On the pseudo-likelihood inference for semiparametric models with boundary problems. *Biometrika* **104** 165–179.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics* **25**, 573–578.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics* **9**, 861–869.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* **101**, 1502–1512.

He, W. and Yi, G. Y. (2011). A pairwise likelihood method for correlated binary data with/without missing observations under generalized partially linear single-index models. *Statistica Sinica* **21**, 207.

Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association* **93**, 1099–1111.

Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika* **90**, 355–366.

Huang, L., Zalkikar, J. and Tiwari, R. C. (2011). A likelihood ratio test based method for signal detection with application to fda's drug safety data. *Journal of the American Statistical Association* **106**, 1230–1241.

Irwig, L., Macaskill, P., Glasziou, P. and Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology* **48**, 119–130.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J. and Thun, M. J. (2009). Cancer Statistics, 2009. *CA: A Cancer Journal for Clinicians* **59**, 225–249.

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–27.

Larribe, F. and Fearnhead, P. (2011). On composite likelihoods in statistical genetics. *Statistica Sinica* **21**, 43.

Liang, K.-Y. (1987). Extended mantel-haenszel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289–299.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.

Lo, J. A. and Fisher, D. E. (2014). The melanoma revolution: from uv carcinogenesis to a new era in therapeutics. *Science* **346**, 945–949.

McVean, G. A., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R. and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.

Moses, L. E., Shapiro, D. and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine* **12**, 1293–1316.

Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324.

Niu, M. T., Erwin, D. E. and Braun, M. M. (2001). Data mining in the us vaccine adverse event reporting system (vaers): early detection of intussusception and other events after rotavirus vaccination. *Vaccine* **19**, 4627–4634.

Nott, D. J. and Rydén, T. (1999). Pairwise likelihood methods for inference in image models. *Biometrika* **86**, 661–676.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). Subsampling Springer-Verlag, New York.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* **58**, 982–990.

Rutter, C. and Gatsonis, C. (1995). Regression methods for meta-analysis of diagnostic test data. *Academic Radiology* **2**, S48–56.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Shimabukuro, T. T., Nguyen, M., Martin, D. and DeStefano, F. (2015). Safety monitoring in the vaccine adverse event reporting system (vaers). *Vaccine* **33**, 4398–4405.

Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 275–296.

Susko, E. (2013). Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom. *Biometrika* **100**, 1019–1023.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

Varin, C. and Vidoni, P. (2008). Pairwise likelihood inference for general state space models. *Econometric Reviews* **28**, 170–185.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes.

Journal of the Royal Statistical Society. Series B (Statistical Methodology) **50**, 297–312.

Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics* **28**, 779–814.

Xing, Y., Bronstein, Y., Ross, M. I., Askew, R. L., Lee, J. E., Gershenwald, J. E., Royal, R. and Cormier, J. N. (2011). Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis. *Journal of the National Cancer Institute* **103**, 129–142.

Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA.

E-mail: jing14@upenn.edu

Department of Statistical Science, Cornell University, Ithaca, New York 14853, U.S.A.

E-mail: yn265@cornell.edu

AT&T Chief Data Office, AT&T Services Inc., Plano, Texas, U.S.A.

E-mail: yc010e@att.com

Institute of Population Health Science, National Health Research Institutes, Zhunan, Miaoli County 35053, R.O.C.

E-mail: kyliang@nhri.org.tw

Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

E-mail: ychen123@upenn.edu

(Received March 2016; accepted July 2018)